

Maryland Council on Open Data Meeting Minutes

100 Community Place, Crownsville, MD 21032

July 20, 2018 | 10 AM – 12 PM

Introductions and Opening Remarks

- A. Mike Leahy, Secretary of the Department of Information Technology and Chairman of the Maryland Council on Open Data, welcomed everyone to the Maryland Council on Open Data meeting.
- B. Patrick McLoughlin provided an overview of the meeting agenda and noted the guidelines are currently under review by legal and will come back for review during the September meeting.
- C. Introductions around the room.

New Item – Data Subcommittee/ Work Group

- A. Patrick McLoughlin discussed the implementation of a subcommittee that is more technically focused. These people would work day-to-day with the data, understanding current data, offering improvements, possessing knowledge of applications being used.
- B. Looking into meeting more regularly. Monthly? Bi-monthly?
- C. Delves into more details then brings recommendations back to the Council.
- D. Subcommittee members do not need to be on the Council.
- E. Looking for interest from meeting attendees or if they know of others at their agencies.
- F. Question – Doesn't DoIT already look into the data?
 - a. DoIT establishes standards. The subcommittee would focus on what data is out there and available.
 - b. Secretary Leahy noted that it is a mistake to think data is a subset of information technology. The process should be: have a business process, come to DoIT with process, find out what data is needed to implement business practice, figure out how to obtain the data, then apply the technology to achieve the business process.
- G. Question – Is there a way to make the data more interesting for researchers to analyze?
 - a. We don't make the data more interesting. We want to make it clear what data is available and that data is in a format that is standardized. We want to make sure the data is not redundant or obsolete.
- H. Patrick will send out an email to meeting attendees to get interest and topics for a subcommittee meeting within the next two months.

New Item – Spring Cleaning Follow-Up & Demo

- A. Patrick McLoughlin stated that for the past 18 months the focus has been to keep data updated and fresh.
- B. Each dataset gets assigned an update frequency schedule, which is how often a dataset owner has to update the data on the Portal.
- C. There is a daily automated report that shows how many datasets are being maintained.
- D. For the past three quarters, under 20 datasets were not being maintained.
- E. Recently, that number has jumped up to 91 datasets. The plan is to go back and make sure we get back on track.
 - a. Make sure DoIT has the appropriate contact at each agency to update the data.
 - b. Ask agencies to look at their data and determine who should maintain it and when it can be updated.
- F. Mike Morello (GOPI) discussed standard operating procedures.
 - a. Urged agencies to make data freshness a priority to get the number back down to 0.
 - b. Informed agencies he has been providing the Governor's Chief of Staff with updates on data freshness every two weeks and will continue to do so until the number of out-of-date datasets drops significantly.
 - c. Hoping there does not need to be an escalation to the Secretary.
 - d. Secretary Leahy noted to find out why agencies stopped updating their data and address those concerns.
- G. If data is being stored in separate places, please let DoIT know so the process can be streamlined.
- H. About 15% radio silence from agencies with outdated datasets. Sometimes one radio silent agency owns multiple datasets.
- I. Question – Why is the crime data from 2016 the most current data available and why is it considered relevant?
 - a. Counties send their data in per month, which then goes to the FBI. The FBI does their review and collation. The aggregated data is then pulled from the FBI, added to the state crime stats and posted to the portal. This is what causes the lag in posting time.
 - b. Some counties have data on their website, which is federated with the Open Portal. We can attempt to federate additional counties crime data into the portal (if available).
 - c. It can be considered relevant when looking at year to year crime statistics
- J. Question – Are there limited number of rows in datasets?
 - a. No.
- K. Plan is to redesign the Open Data Portal.

Data Review – Open Data Portal & GIS Portal Review & Stats

- A. Growth in both portals.
- B. Input received notes the portals are useful and beneficial, which puts responsibility on us to keep data fresh and clean.
- C. Total user of Open Data Portal: 15,589; total users of GIS Portal: 11,479
- D. Bounce rate for Open Data Portal: 59%; bounce rate for GIS Portal: 35%. DoIT is looking at why the bounce rate is so high – would like it to be lower.
- E. Question – What is bounce rate?
 - a. The percentage of single page visits. When a user leaves the portal from the landing page without browsing any further
- F. Over 14,000 new users in Open Data Portal; over 10,000 new users in GIS Portal. Most users are from outside the State of Maryland.
- G. Current number of datasets: 343 in Open Data Portal; 917 in GIS Portal.
- H. Review of most visited pages.
- I. Number of separate session page visits:
 - a. **The Behavioral flow of users shows a 70% drop off after the user has visited the dataset page, a 38% drop off of those users after they have navigated to a second dataset page, a 47% drop off of the remaining users after visiting a third dataset page, and a 32% drop off of users who navigate to a fourth dataset**
- J. Question – How do agencies decide what to publish?
 - a. One on one conversations with the agencies to decide.
- K. Question – How does one put in a request for data?
 - a. Requests should be sent to Patrick McLoughlin.
- L. Direct and organic searches are the primary source of traffic on Portals. Referrals are the second source of traffic on Portals. Patrick McLoughlin will provide additional statistics as part of meeting minutes:
 - a. **53.7% through organic search (google or similar search engine search)**
 - b. **39.5% through direct search (direct input of data.maryland.gov)**
 - c. **6.5% through referrals (linked in from other website)**
 - d. **0.3% through social media**
- M. Patrick McLoughlin will provide additional stats as part of meeting minutes:
 - a. **503,000 dataset API calls in the month of May.**
- N. Working with Socrata to silo data and make sure there are no duplicates.

Data Review – New Features & Datasets

- A. Patrick McLoughlin noted the new features are being put into place for ease of access.
- B. Search Portal/ Bar

- a. Ability to search keywords and get results from all Maryland datasets.
 - b. Suggestion – Tag of agency that owns the dataset
 - c. Suggestion – Add description of dataset to search results
 - d. Question – Will there be standards for how people title their datasets?
 - i. Yes.
 - e. Timeline: next week.
- C. Data Inspector/ Profiler
- a. Feature will be used to get a better handle on what our data looks like, how often the data is updated, etc.
 - b. Need more information about how much is changed when a dataset is updated.
 - c. Scripts will go through each cell of datasets.
 - d. There will be a search capability.
 - e. Data in table can be exported.
 - f. Ability to take action based on what results we are seeing.
 - g. Timeline: go live on Monday (July 23)
- D. Data Owner Update Alerts
- a. Dataset owners will get email alerts when it is time to update the data.
 - b. Working on setting standards regarding when alerts should go out based on the different types of data.
 - c. Want to give owners enough time to update the data but don't want to send it so far out that it gets ignored and forgotten.
 - d. Want to coordinate so DoIT is aware of other contacts helping agencies update their datasets.
 - e. Hoping that this feature helps ensure we do not need to escalate anything to the Secretary's level.
 - f. Question – Would it make sense to add Council members to the email alerts so they are aware?
 - i. Right now, only the dataset owner gets the email alert; however, we can add anyone to the email alerts.

Open Discussion

- A. Secretary Leahy's data standards document will be distributed. Plan on discussing next meeting.
- B. Assistant Attorney Generals are looking into GDPR (General Data Protection Regulation) not just effecting international unit.
- C. Summary Stats for next meeting. Going to revisit the idea of looking at the kinds of information people request in PIA requests.
- D. Aim for 1,500 datasets in the Portal. Look at what data from agencies can be added into the Portal throughout the next quarter.

- E. Mike Morello to go to agencies that had no attendee to reaffirm their commitment to the Council on Open Data.
- F. Possibility to list agencies that were not in attendance on home page?
- G. Idea to use “how-to” videos to walk users through features on the Portals.
- H. Idea to add a “people who searched X topic also searched X topic” feature to Portals.
- I. Idea to add a tab on website for videos/ resources.

Closing

Meeting was adjourned.

Next meeting is September 21, 2018 from 10:00 am – 12:00 pm.